AFRL-AFOSR-VA-TR-2015-0267

SOCIO CULTURAL MEDIA SHARING AS CONVERSATIONS SENSING AND MODELING BEHAVIOR IN RESPONSE to Environmental Changes

Michael Stefanone
RESEARCH FOUNDATION OF STATE UNIVERSITY OF NEW YORK THE

07/06/2015
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
05/12/2015	Final Progress	04-01-12 to 03-31-15

4. TITLE AND SUBTITLE
Socio Cultural Media Sharing as Conversations Sensing and Modeling Behavior in Response to Environmental Changes

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0201

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Stefanone, Michael; Fu, Yun

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Research Foundation of State University of New York
Sponsored Projects Services
402 Crofts Hall, Univ. at Buffalo
Buffalo, NY 14260

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 N. Randolph St.
Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION A

13. SUPPLEMENTARY NOTES
N/A

14. ABSTRACT
Over the course of this project, the team worked on several fronts to tackle problems related to evaluating and validating the machine learning techniques that were being developed to classify content shared via social media. We began by focusing on text-based content shared via social media. Here we collected data in the form of social media user's comments to news articles about hydrofracking (or, fracking). This topic was chosen because it is a timely social and political issue.

15. SUBJECT TERMS
Machine learning, social media

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Michael Stefanone
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	4	716-645-1159

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATE COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33315-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report. e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Summary: Over the course of this project, the team worked on several fronts to tackle problems related to evaluating and validating the machine learning techniques that were being developed to classify content shared via social media. We began by focusing on text-based content shared via social media. Here we collected data in the form of social media user's comments to news articles about hydrofracking (or, *fracking*). This topic was chosen because it is a timely social and political issue. Nearly 300 posts were analyzed with automated linguistic inquiry tools and results show that systematic patterns in language use were predicted by geographic location of users. Next, we expanded our efforts to address image-based content shared via social media. We opted to collect all images shared via twitter associated with the active conversation on the topic of gun control (*#guncontrol*). Again, this topic was chosen given it's central position in the public sphere concerning social and policy issues in the United States. Our team first wrote a custom python script to track and collect all media—specifically images—shared over the course of one month on twitter, via the conversation identified by #guncontrol. We simultaneously developed a human image coding protocol so that shared images could be classified along dimensions including frame, appeal, and valence. This classification protocol was designed so that it could be used by the machine learning algorithms. Next we trained human coders, and classified all images. We were able to predict which image attributes predicted diffusion and propagation across twitter. However, problems included the difficultly of obtaining consistent results from human coders, because human coders vary in terms of their positions on political and social issues like gun control. Results indicate that shared images with attribute frames, fear and humor appeals, and positive valence are retweeted more often. Also retweeted more frequently are messages from users with larger networks and whose tweets contain hashtags. Results also show a significant negative relationship between the time since the last major shooting event in the United States and the likelihood that messages with images are retweeted. These results are meaningful considering the context of evolving mass media systems and online social networks.

Media Coverage vs. Online Discussion: We aim to compare the public opinions of specific topics like "Gun Control" on media coverage and online discussion. The LexisNexis database provides archived newspaper articles and news transcripts from major TV networks, which can be used as the media coverage part. For the online discussion part, we developed software to crawl comments from Facebook pages of mass media (e.g. CNN News on Facebook) and comments on mass media's websites. By specifying the date range, our software is able to crawl the online comments automatically. In total, we have collected over 1 million comments from Facebook pages of mass media including CNN, ABC, CBC, NBC News, Fox, WSJ, etc. Some content analysis software (e.g., LIWC) could be used to analyze the crawled data. To facilitate the analysis, we also developed several text processing tools to deal with the output results of LIWC. We developed a web application (crawler), which fetch news from NY Times. It's

developed using the API provided by NY Times. Our crawler has a simple GUI that allows users to search news about certain keywords in a certain period, and saves the results in the commonly used CSV format. A further study is an exploratory attempt to use automatic linguistic analysis for understanding social media users' news commenting behavior. The study addresses geographically–based dynamics in human–computer interaction, namely, users' tie to a geographic community. Specifically, the study reveals that commenting behavior differs between users of different levels of local community tie. Comments by local users, those with higher level of local community tie, exhibit different linguistic patterns in comparison to national users who are less involved in local community. The linguistic differences are reflected in the use of pronouns, personal pronouns, social words, swear words, anxiety words and anger words. We argue that identification of these differences is crucial in the practice of mining social media conversations for public opinion.

Image Retweeting: To study the image retweeting problem, we developed a web application that fetches tweets from twitter with given keywords, in a certain time period. It's developed based on the API of Twitter. In order to facilitate our analysis, we applied machine learning and computer vision algorithms to the crawled images. One interesting application is human identification. Given an input image, we need to determine whether it contains a human body. Particularly, we utilized a part-based human detection algorithm on the collected image data set. Some images in this data set have been manually labeled by our collaborators, which enables us to evaluate the performance of machine learning algorithms. Subspace segmentation is one of the hottest issues in computer vision and machine learning fields. Generally, data (e.g., images of human faces) are lying in a union of multiple linear subspaces, therefore, it is the key to find a block diagonal affinity matrix, which would result in segmenting data into different clusters correctly. Recently, graph construction based segmentation methods attract lots of attention. Following this line, we propose a novel approach to construct a Sparse Graph with

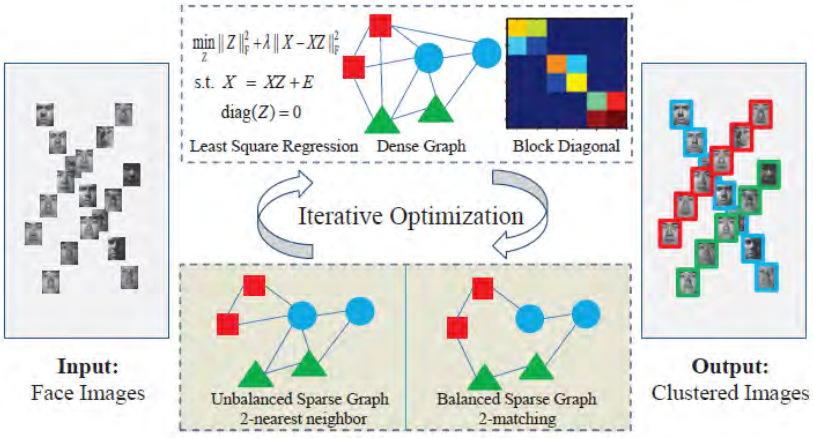

Fig. 1. Framework of our proposed method. Inputs are various facial images taken under different conditions (illumination, expression, gender, etc). Here we show face images from three different individuals. We propose a novel approach to construct a graph (middle part). By virtue of the block diagonal property by Frobenius norm (top-middle) and sparse property (bottom-middle), a block-wise and sparse graph can be achieved. Specifically, in sparsity part, we take both unbalanced data (left) and balance data (right) into consideration using k-nearest neighbor and b-matching method, respectively. The final clustering result is shown in the right part.

Block-wise constraint for face representation, named SGB (Fig. 1). Inspired by the recent study of least square regression coefficients, SGB firstly generates a compact block-diagonal coefficient matrix. Meanwhile, graph regularizer brings in a sparse graph, which focuses on the local structure and benefits multiple subspaces segmentation. By introducing different graph regularizers, our graph would be more balanced with b-matching constraint for balanced data. By using knearest neighbor regularizer, more manifold information can be preserved for unbalanced data. To solve our model, we come up with a joint optimization strategy to learn block-wise and sparse graph simultaneously. To demonstrate the effectiveness of our method, we consider two application scenarios, i.e., face clustering and kinship verification. Extensive results on Extended YaleB and ORL demonstrate that our graph consistently outperforms several state-of-the-art graphs. Our algorithm achieves an accuracy of more than 98% (see following table). We also developed algorithms to automatically analyze the facial expressions, age, and emotions.

AVERAGE CLUSTERING ACCURACY (%) OF DIFFERENT METHODS ON DATASET ORL AND EXTENDED YALEB (5 AND 10 SUBJECTS).

Methods	SSC [3]	LRR [12]	LSR [14]	LRCB [10]	CASS [13]	SGB-I	SGB-II
ORL	78.50	77.50	76.80	86.22	81.25	88.60	**89.50**
YaleB-5	80.31	86.56	92.19	95.34	94.03	96.47	**98.13**
YaleB-10	52.19	65.00	73.59	85.45	81.88	92.84	**94.53**

Publication: We have published more than 20 papers in the last 2 years supported by this grant, including 2 best paper awards.

1. Stefanone, M. A., Saxton, G., *Egnoto, M. A., *Wei, W. X., & Fu, R. (2015). Image attributes and diffusion via twitter: The case of#guncontrol(249 KB). In the Proceedings of the 48th Hawaii International Conference on System Sciences (HICSS, 2015), 1788-1798. January 5-8, Kauai.
2. *Xu, W., *Li, L., Stefanone, M. A., & Fu, R. (2014). Does social media users' commenting behavior differ by their local community tie? A computer–assisted linguistic analysis approach. First Monday, 19(1). doi: http://dx.doi.org/10.5210%2Ffm.v19i1.4821
3. Handong Zhao, Zhengming Ding, and Yun Fu, Block-wise Constrained Sparse Graph for Face Image Representation, IEEE Conference on Automatic Face and Gesture Recognition (FG), 2015
4. Ming Shao, Zhengming Ding, and Yun Fu, Sparse Low-Rank Fusion based Deep Features for Missing Modality Face Recognition, IEEE Conference on Automatic Face and Gesture Recognition (FG), 2015
5. Sheng Li, Ming Shao, and Yun Fu, Multi-view Low-Rank Analysis for Outlier Detection, SIAM International Conference on Data Mining (SDM), 2015

6. Sheng Li and Yun Fu, Robust Subspace Discovery through Supervised Low-Rank Constraints, SIAM International Conference on Data Mining (SDM), 2014 (Oral Presentation [**Best Paper Award**] 1 out of 384)
7. Zhengming Ding, Ming Shao, and Yun Fu, Latent Low-Rank Bi-Directional Transfer Subspace Learning, Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI), 2014
8. Shuyang Wang, Jinzheng Sha, Huaiyu Wu, and Yun Fu, HIERARCHICAL FACIAL EXPRESSION ANIMATION BY MOTION CAPTURE DATA, IEEE International Conference on Multimedia and Expo (ICME), 2014
9. Ming Shao, Sheng Li, Tongliang Liu, Dacheng Tao, Thomas S. Huang, and Yun Fu, LEARNING RELATIVE FEATURES THROUGH ADAPTIVE POOLING FOR IMAGE CLASSIFICATION,IEEE International Conference on Multimedia and Expo (ICME), 2014 (Oral Presentation [**Best Paper Award** Candidate] 4 out of 716)
10. Dmitry Kit, Yu Kong and Yun Fu, Location Aware Self-Organizing Map for Discovering Similar and Unique Visual Features of Geographical Locations, International Joint Conference on Neural Networks (IJCNN), 2014
11. Shuyang Wang, Ming Shao and Yun Fu, Attractive or Not? Beauty Prediction with Attractiveness-Aware Encoders and Robust Late Fusion, ACM Multimedia (MM), 2014
12. Zhengming Ding and Yun Fu, Low-Rank Common Subspace for Multi-view Learning, IEEE International Conference on Data Mining (ICDM), 2014 [acceptance rate 9.7%]
13. Book: Yun Fu, Human-Centered Social Media Analytics, Springer, 2014
14. Liangyue Li, Sheng Li and Yun Fu, Learning Low-Rank and Discriminative Dictionary for Image Classification, Image and Vision Computing (IVC), 2014. (Accepted, invited for special issue on The Best of Face and Gesture 2013)
15. Kang Li and Yun Fu, Prediction of Human Activity by Discovering Temporal Sequence Patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), Volume:, Issue:, Page(s):, 2014.
16. Ming Shao, Dmitry Kit, and Yun Fu, Generalized Transfer Subspace Learning through Low-Rank Constraint, International Journal of Computer Vision (IJCV), Volume:, Issue:, Page(s):, 2014.
17. Ming Shao, Liangyue Li, and Yun Fu, What Do You Do? Recognize Occupations in a Photo via Social Context, International Conference on Computer Vision (ICCV), 2013
18. Xu Zhao, Yuncai Liu and Yun Fu, Exploring Discriminative Pose Sub-Patterns for Effective Action Classification, ACM Multimedia(MM), 2013 [Full Paper]
19. Ming Shao, Liangyue Li, and Yun Fu, Predicting Professions through Probabilistic Model under Social Context, AAAI Conference on Artificial Intelligence (AAAI), 2013
20. Sheng Li and Yun Fu, Low-Rank Coding with b-Matching Constraint for Semi-supervised Classification, International Joint Conference on Artificial Intelligence (IJCAI), 2013

AFOSR Deliverables Submission Survey

Response ID:4693 Data

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

bradley.bermudez@buffalo.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

716-645-4383

Organization / Institution name

The Research Foundation for SUNY

Grant/Contract Title
The full title of the funded effort.

Socio Cultural Media Sharing as Conversations Sensing and Modeling Behavior in Response to Environmental Changes

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0201

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Michael A. Stefanone

Program Manager
The AFOSR Program Manager currently assigned to the award

Benjamin Knott

Reporting Period Start Date

04/01/2012

Reporting Period End Date

03/31/2015

Abstract

The project team worked on several fronts to tackle problems related to evaluating and validating the machine learning techniques that were being developed to classify content shared via social media. We began by focusing on text-based content shared via social media. Here we collected data in the form of social media user's comments to news articles about hydrofracking (or, fracking). Next, we expanded our efforts to address image-based content shared via social media. We opted to collect all images shared via twitter associated with the active conversation on the topic of gun control (#guncontrol). We simultaneously developed a human image coding protocol so that shared images could be classified along dimensions including frame, appeal, and valence. Next we trained human coders, and classified all images. We were able to predict which image attributes predicted diffusion and propagation across twitter. However, problems included the difficultly of obtaining consistent results from human coders, because human coders vary in terms of their positions on political and social

issues like gun control. Results indicate that shared images with attribute frames, fear and humor appeals, and positive valence are retweeted more often. Also retweeted more frequently are messages from users with larger networks and whose tweets contain hashtags. Results also show a significant negative relationship between the time since the last major shooting event in the United States and the likelihood that messages with images are retweeted. These results are meaningful considering the context of evolving mass media systems and online social networks.

Distribution Statement
This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement
If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form
Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

Stefanone AFOSR SF 298.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

Stefanone AFOSR finalReport.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

Changes in research objectives (if any):

Change in AFOSR Program Manager, if any:

Extensions granted or milestones slipped, if any:

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Jun 25, 2015 13:44:36 Success: Email Sent to: bradley.bermudez@buffalo.edu